



UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/823,052	03/30/2001	Yevgeny Agichtein	14367	8988
23389	7590	10/29/2004	EXAMINER	
SCULLY SCOTT MURPHY & PRESSER, PC 400 GARDEN CITY PLAZA GARDEN CITY, NY 11530			WOZNIAK, JAMES S	
			ART UNIT	PAPER NUMBER
			2655	

DATE MAILED: 10/29/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

Office Action Summary

Application No.

09/823,052

Applicant(s)

AGICHTEIN ET AL.

Examiner

James S. Wozniak

Art Unit

2655

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

- 1) ☒ Responsive to communication(s) filed on 3/30/2001.
- 2a) ☐ This action is **FINAL**. 2b) ☒ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

- 4) ☒ Claim(s) 1-60 is/are pending in the application.
- 4a) Of the above claim(s) _____ is/are withdrawn from consideration.
- 5) ☐ Claim(s) _____ is/are allowed.
- 6) ☒ Claim(s) 1-60 is/are rejected.
- 7) ☐ Claim(s) _____ is/are objected to.
- 8) ☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on _____ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some * c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
2. ☐ Certified copies of the priority documents have been received in Application No. _____.
3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

- 1) ☐ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08)
Paper No(s)/Mail Date 3/30/2001.
- 4) ☐ Interview Summary (PTO-413)
Paper No(s)/Mail Date. _____.
- 5) ☐ Notice of Informal Patent Application (PTO-152)
- 6) ☐ Other: _____.

DETAILED ACTION

Claim Rejections - 35 USC § 112

1. **Claims 41 and 42** are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

2. **Claims 41 and 42** recite the limitation "the query" in line 2 of each claim.

There is insufficient antecedent basis for this limitation in the claims.

Claim Rejections - 35 USC § 102

3. The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the basis for the rejections under this section made in this Office action:

A person shall be entitled to a patent unless –

(b) the invention was patented or described in a printed publication in this or a foreign country or in public use or on sale in this country, more than one year prior to the date of application for patent in the United States.

4. **Claims 1-4, 6-15, 17-24, 26-30, 36-50, and 52-60** are rejected under 35 U.S.C. 102(b) as being anticipated by Kupiec (*U.S. Patent: 5,696,962*).

With respect to **Claim 1**, Kupiec discloses:

Generating a set of phrases that identify different categories of questions (*Col. 7, Line 63-Col. 8, Line 3*);

Art Unit: 2655

Generating candidate transforms for each phrase (*result pairs including query terms, Col 22, Line 64- Col. 24, Line 15*);

Weighting the candidate transforms (*Col. 24, Lines 40-42*);

Ranking the candidate transforms (*Col. 24, Lines 16- Col. 25, Line 18*);

Applying the transforms to an information retrieval system (*using ranked result pairs to create a presentation list, Col. 25, Lines 20-33*).

With respect to **Claim 2**, Kupiec recites:

Filtering candidate transforms prior to weighting (*Col. 24, Lines 26-35*).

With respect to **Claim 3**, Kupiec discloses:

Natural language processing techniques are used for filtering (*Col. 8, Lines 46-61*).

With respect to **Claim 4**, Kupiec discloses:

A natural processing technique used is part-of-speech tagging (*Col. 8, Lines 46-61*).

With respect to **Claim 6**, Kupiec recites:

The natural language processing techniques used are feature selection techniques (*Col. 9, Lines 34-45*).

With respect to **Claim 7**, Kupiec recites:

The questions are categorized by similar goals (*query relationships, Col. 10, Line 29- Col. 11, Line 13*).

With respect to **Claim 8**, Kupiec discloses:

The categories are identified using an n-gram approach (*n-tuple proper noun- New York City borough, Col. 9, Line 10- Col. 10, Line 8*).

With respect to **Claim 9**, Kupiec recites:

Phrases are generated by computing the frequency of all n-grams of length minQtokens to maxQtokens words, with all n-grams anchored at the beginning of the questions (*frequency table, Col. 13, Line 1- Col. 14, Line 41*).

With respect to **Claim 10**, Kupiec recites:

All n-grams that occur at least minQphraseCount times are used for generating question phrases (*Col. 15, Lines 1-7, and Col. 34, Lines 1-9*).

With respect to **Claim 11**, Kupiec recites:

The input to generating the set of phrases is a set of questions (*Col. 5, Lines 51-58*).

With respect to **Claim 12**, Kupiec recites:

The output to generating the set of phrases is a set of question phrases that can be used to classify questions into respective question types (*noun and title phrases, Col. 8, Line 46- Col. 10, Line 8*).

With respect to **Claims 13-15**, Kupiec discloses:

Filtering the generated phrases using a natural language part-of-speech tagging technique (*Col. 8, Lines 46-61*).

With respect to **Claim 17**, Kupiec recites:

The natural language processing techniques used are feature selection techniques (*Col. 9, Lines 10-45*).

With respect to **Claim 18**, Kupiec discloses:

Generating candidate transforms comprises generating initial candidate transform phrases (*unprocessed result pairs, Col. 8, Lines 16-17, and Col. 23, Lines 44-48*).

With respect to **Claim 19**, Kupiec recites:

Initial candidate transforms are generated by using a collection of Question/Answer pairs
(*Col. 23, Lines 44-48*).

With respect to **Claim 20**, Kupiec discloses:

Filtering initial candidate transform phrases (*Col. 24, Lines 26-35*).

With respect to **Claim 21**, Kupiec discloses:

Initial candidate transforms are filtered by minimum co-occurrence (*Col. 34, Lines 1-9*).

With respect to **Claim 22**, Kupiec discloses:

Weighting filtered initial candidate transforms (*Col. 24, Lines 40-42*).

With respect to **Claim 23**, Kupiec discloses:

Filtering all weighted initial candidate transforms (*third filtering criterion, Col. 24, Lines 45-52*).

With respect to **Claim 24**, Kupiec discloses:

The collection of pairs has been tagged with a part-of-speech tagger (*Col. 8, Lines 46-61*).

With respect to **Claim 26**, Kupiec recites:

Initial candidate transform phrases are filtered by eliminating generated answer phrases that contain a noun (*Col. 9, Lines 19-29*).

With respect to **Claim 27**, Kupiec discloses:

All potential answer phrases are generated from all of the words in the prefix of Answer for each Question/Answer pair where a prefix of Question matches each question phrase (*Col. 31, Lines 11-33, and Col. 36, Lines 9-15*).

With respect to **Claim 28**, Kupiec recites:

N-grams of length minAtokens to maxAtokens words are used, starting at every word boundary in the first maxLen bytes of the Answer text (*Col. 31, Lines 11-33, and Hoover Dam, Col. 36, Lines 9-15*).

With respect to **Claim 29**, Kupiec recites:

From the resulting n-grams, the topKphrases with the highest frequency counts are kept (*Col. 31, Lines 1-10*).

With respect to **Claim 30**, Kupiec discloses:

Information retrieval techniques for term weighting is applied to rank the initial candidate transforms (*number of proper noun matches in retrieved documents, Col. 24, Lines 35-45*).

With respect to **Claim 36**, Kupiec recites:

Sorting the initial candidate transforms into buckets according to the number of words in the transform phrase, and up to maxBucket transforms, with the highest values of term selection weights kept from each bucket (*grouping and ranking result pairs according to the number of title words, Col. 24, Lines 35-44, and the determination of the most relevant result pairs for a presentation list, Col. 25, Lines 20-33*).

With respect to **Claim 37**, Kupiec discloses:

Filtering and weighting the initial candidate transform prior to sorting (*Col. 10, Lines 9-29, and Col. 24, Lines 36-44*).

With respect to **Claim 38**, Kupiec discloses:

Ranking the candidate transforms comprises retrieving a set of Question/Answer pairs (*Col. 23, Line 60- Col. 24, Line 15*) and for each pair and the candidate transforms, applying a transform to each Question (*Col. 23, Lines 7-35 and Col. 59-65*).

Art Unit: 2655

With respect to **Claim 39**, Kupiec discloses:

Sorting Question/Answer pairs by increasing answer length prior to ranking the candidate transforms (*Col. 24, Lines 45-52*).

With respect to **Claim 40**, Kupiec recites:

The transforms are encoded so that they are treated as phrases by the information retrieval system (*Col. 31, Lines 25-45*).

With respect to **Claim 41**, Kupiec recites:

Question requires parts of the query in matching pages (*Col. 32, Lines 19-22*).

With respect to **Claim 42**, Kupiec discloses:

Question does not require parts of the query in matching pages (*discarding query portions, Col. 9, Lines 24-33*).

With respect to **Claim 43**, Kupiec recites:

Multiple transformations are combined into a single query (*Col. 35, Lines 1-5*).

With respect to **Claim 44**, Kupiec recites:

Categorizing questions asked of the information retrieval system into different types (*type phrase, Col. 5, Lines 15-24, and Col. 29, Lines 19-53*);

Generating phrases that identify the question types (*Col. 9, Line 48- Col. 10, Line 8*);

Generating candidate query transformations for each phrase from a training set of question/answer pairs (*result pairs, Col. 23, Lines 44- Col. 24, Line 15*);

Evaluating the candidate transforms on the target information retrieval systems (*Col. 24, Lines 16- Col. 25, Line 18*), and;

Applying transformations to queries submitted to the information retrieval system
(*related words, Col. 23, Lines 7-35*).

With respect to **Claim 45**, Kupiec recites:

The questions are categorized by similar goals (*query relationships, Col. 10, Line 29-
Col. 11, Line 13*)

Claim 46 contains subject mater similar to Claim 9, and thus, is rejected for the same reasons.

Claim 47 contains subject mater similar to Claim 10, and thus, is rejected for the same reasons.

Claims 48-50 contain subject matter similar to Claims 13-15, respectively, and thus, are rejected for the same reasons.

Claim 52 contains subject mater similar to Claim 17, and thus, is rejected for the same reasons.

With respect to **Claim 53**, Kupiec discloses:

Filtering, weighting, and ranking the candidate query transformations prior to evaluating on the information retrieval systems (*Col. 24, Line 16- Col. 25, Line 34*).

With respect to **Claim 54**, Kupiec recites:

Natural language processing techniques are used for filtering (*Col. 24, Lines 26-44*).

Claim 55 contains subject mater similar to Claim 21, and thus, is rejected for the same reasons.

Claim 56 contains subject mater similar to Claim 18, and thus, is rejected for the same reasons.

Claim 57 contains subject matter similar to Claim 20, and thus, is rejected for the same reasons.

Claim 58 contains subject matter similar to Claim 4, and thus, is rejected for the same reasons.

Claim 59 contains subject matter similar to Claim 16, and thus, is rejected for the same reasons.

With respect to **Claim 60**, Kupiec discloses:

Entering a question whose answer is desired (*Col. 28, Lines 54-67*);

Classifying the question by matching it with predetermined question phrases (*type phrase, Col. 29, Lines 19-53*);

Retrieving the associated question phrases (*Col. 31, Lines 25-33*);

Rewriting the question by applying each associated question phrase to the question to create transformed queries (*alternative interpretations, Col. 31, Lines 1-10*);

Submitting the transformed queries to an information retrieval system (*Col. 31, Lines 1-33*);

Analyzing the returned documents (*Col. 31, Lines 34-36*);

Scoring the returned documents (*Col. 31, Lines 34-36*);

Ranking the returned documents by their respective scores (*Col. 31, Lines 40-41*), and;

Documents ranked above a predetermined level as the resulting retrieved documents (*subset of most relevant ranked documents that would require an inherent threshold level for determining relevancy, Col. 31, Lines 40-45*).

Claim Rejections - 35 USC § 103

5. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

6. **Claims 5, 16, 25, and 51** are rejected under 35 U.S.C. 103(a) as being unpatentable over Kupiec.

With respect to **Claims 5, 16, 25, and 51**, Kupiec teaches the natural language document retrieval system and method utilizing part-of-speech tagging, as applied to Claims 4, 15, 24, and 50. Although Kupiec does not specifically suggest the use of a Brill's part-of-speech tagger, the examiner takes official notice that it would have been obvious to one of ordinary skill in the art, at the time of invention, to utilize a Brill's part-of-speech tagger for the tagger disclosed by Kupiec since Brill's part-of-speech tagger is well-known in the art, common in usage, and readily available. Therefore, in order to implement a part-of-speech tagger that is readily available, it would have been obvious to one of ordinary skill in the art, at the time of invention, to modify the teachings of Kupiec with a Brill's part-of-speech tagger.

7. **Claims 31-34** are rejected under 35 U.S.C. 103(a) as being unpatentable over Kupiec in view of Robertson et al ("*Relevance weighting of Search Terms*," 1976).

With respect to **Claim 31**, Kupiec teaches the natural language document retrieval system utilizing term weighting, as applied to Claim 30. Kupiec does not specifically suggest using a

Art Unit: 2655

Sparck Jones weighting scheme, however, such a scheme is well known in the art, as is evidenced by Robertson (*Page 129, Col. 2–130, Col. 1*).

Kupiec and Robertson are analogous art because they are from a similar field of endeavor in document retrieval systems. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to modify the teachings of Kupiec with the Sparck Jones weighting scheme to improve document retrieval accuracy and, in turn, retrieval performance (*Robertson, Page 129, Col. 2–130, Col. 1*).

With respect to **Claim 32**, Robertson additionally discloses:

Candidate transforms are weighted by assigning to each phrase a Robertson/Sparck Jones term weight with respect to a specific question type (*term weighting utilizing relevance to user strategy, Page 131*).

With respect to **Claim 33**, Robertson additionally discloses:

The weight is computed for each candidate transform $tr.sub.i$ by computing the count of Question/Answer pairs where $tr.sub.i$ appears in the Answer to a question matching a question phrase as the number of relevant documents (*relevance weighting, Page 131*);

Considering the number of remaining Question/Answer pairs where $tr.sub.i$ appears in the Answer as non-relevant (*non-relevance odds, Page 131, Col. 2*) and;

Applying the formula $5 w_i(1) = (r + 0.5) / (R - r + 0.5) (n - r0.5) / (N - n - R + r + 0.5)$ (*weighting function, Page 130, and further derivation using the document relevance table on Page 135*).

With respect to **Claim 34**, Robertson additionally discloses:

Term selection weights are computed for each candidate transform (*weighting each search term, Page 131*).

8. **Claim 35** is rejected under 35 U.S.C. 103(a) as being unpatentable over Kupiec in view of Robertson et al (*Okapi at TREC-7: Automatic Ad Hoc, Filtering, VLC, and Interactive track, " 1998*).

With respect to **Claim 35**, Kupiec teaches the natural language document retrieval system utilizing term weighting, as applied to Claim 30. Kupiec does not specifically suggest a specific term selection weight calculation as recited in Claim 35, however Robertson discloses:

Term selection weights, $w_{tr.sub.i}$, for each candidate transform $tr.sub.i$, are computed as: $w_{tr.sub.i} = q_{tf.sub.i} \cdot w_{sub.i}^{sup.(1)}$ where $q_{tf.sub.i}$ is the co-occurrence count of $tr.sub.i$ with QP, and $w_{sub.i}^{sup.(1)}$ is the relevance-based term weight of $tr.sub.i$ computed with respect to QP (*term ranking for selection, Page 3*).

Kupiec and Robertson are analogous art because they are from a similar field of endeavor in document retrieval systems. Thus, it would have been obvious to a person of ordinary skill in the art, at the time of invention, to modify the teachings of Kupiec with the aforementioned formula for calculating term selection weights as taught by Robertson in order to provide a well known and effective means of determining a query term ranking (*Robertson, term ranking for selection, Page 3*) for further identifying a most relevant search term, thus implementing a more effective document retrieval method.

Conclusion

9. The prior art made of record and not relied upon is considered pertinent to applicant's disclosure:

- Schultz (*U.S. Patent: 5,737,734*)- teaches a method for adjusting query term relevance.
- Noguchi et al (*U.S. Patent: 5,991,755*)- teaches the use of a weighting factor for search terms in a document retrieval system.
- Nelson et al (*U.S. Patent: 5,937,422*)- teaches a means of assigning weighted and ranked search terms to a document.
- Lamburt (*U.S. Patent: 6,374,241*)- teaches a method of query classification.
- Li (*U.S. Patent: 6,480,743*)- teaches a method of query expansion.


10. Any inquiry concerning this communication or earlier communications from the examiner should be directed to James S. Wozniak whose telephone number is (703) 305-8669 and email is James.Wozniak@uspto.gov. The examiner can normally be reached on Mondays-Fridays, 8:30-4:30.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Doris To can be reached at (703) 305-4827. The fax/phone number for the Technology Center 2600 where this application is assigned is (703) 872-9306.

Art Unit: 2655

Any inquiry of a general nature or relating to the status of this application or proceeding should be directed to the technology center receptionist whose telephone number is (703) 306-0377.

James S. Wozniak
10/14/2004


SUSAN MCFADDEN
PRIMARY EXAMINER